7

Audio Data

Audio data is primarily produced by hydrophones (underwater microphones). ONC hosts a large number of hydrophones (over 160 devices, 11 different types) that are the single largest contributor to the archive (about 70% of the 1.2 petabytes of storage). These hydrophones produce audio files with sounds at a wide range of frequencies, having applications in seismology, marine mammal studies, ship noise and more.

> ⓘ Audio data may be diverted or delayed for security reasons or bandwidth limitations. Read on below for more information.

> ⓘ We have recently switched our primary storage format for Audio Data from compressed wav files to FLAC files. FLAC is more space and CPU efficient, and is widely supported. As of April 9, 2021, all hydrophones produce FLAC files as source. Prior to that, there was a transition period when FLAC files were generated from the source wav files and archived as well, that period was August 10, 2020 to April 9, 2021. During this transition period, both wav and FLAC formats were archived. File formats that are archived are readily available, while other formats are generated on-the-fly and are much slower to access. For fast results for lossless audio data, **select wav file formats up to August 10, 2020, and select FLAC thereafter**. The data availability graph in Data Search maybe updated in the future to show which files are available directly, while the archivefiles service can offer a list of files in the archive as well.
>
> If users create large requests that generate FLAC/MP3/WAV formats on-the-fly, the temporary space to hold these data products may fill up and **their search requests will be cancelled and the data deleted**. Searches may also be cancelled for very large search bounds if the search will take too long complete without being interrupted by semi-monthly software maintenance. Please see the above date ranges to create searches that only retrieve data from the archive and avoid generating data products on-the-fly. While the temporary / holding space is quite large, it is easy to request over 10 terabytes of data in a single search. We're working on a more permanent fix for this issue. Please accept our apologies if your search requests are impacted. Cancellations will only be done if necessary and the user will be contacted via email.
>
> Another very good workaround is to request data in smaller "chunks" of time. Users can also do this programmatically using the dataProductDelivery Service (Oceans 3.0 API), so that users can download and process the data as they go. Another alternative is to run processing code or software on the ONC cloud. This avoids downloading the data entirely, removing the limitations of local computing resources. Contact us if you're interested.

> ⓘ On-the-fly conversion between audio formats is now disabled by default because of the tendency to fill the temporary space as noted above. With the default option, searches may return no data found for some formats while there is data in other audio formats. FLAC and wav formats are available as noted above. MP3 format is not recommended and is almost always generated on-the-fly. See Audio Format Conversion for more information.
>
> IOS Hydrophone Arrays always convert on-the-fly from their source raw data .hyd files to audio file products - this is slow and may also fill the temporary space. Contact us before requesting IOS Hydrophone Array audio data.

Given the sensitive nature of hydrophone data, the military has the ability to divert the audio data as required. Diverted data is then reviewed by military authorities, and if it does not contain sensitive recordings it is returned to the ONC archive. This process usually blocks out a few days of near-live data and then the review and return takes up to 3 months (usually about a month). Diversion occurs securely at the source and ONC has no access to the data until it is reviewed and returned. The data products, Data Preview and data availability are all updated automatically when the data is returned. Because hydrophones generate multiple types of data, the data availability plot in Data Search may show available data when audio data is not available. This often indicated by the tones of the colour in the data availability plot:
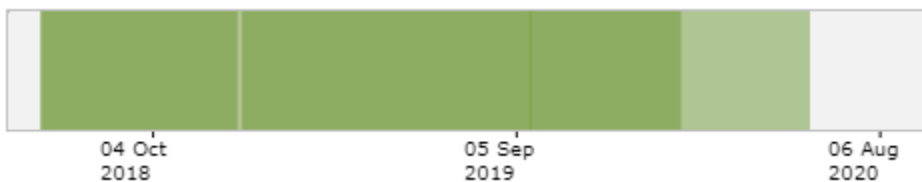

(Light green means less data, likely that the audio data is not available). The other data provided by hydrophones are scalar sensor data, stored as log .txt files, and FFT files, both of which are not subject to diversion. Spectrogram plots and associated data products can be made from either FFT file or audio data as source files or a combination of both; users will see the option to choose the source in a few places (Data Search and Search Hydrophone Data pages). Data Preview's monthly tab shows both the FFT source and audio source spectrograms for the past month - users can also refer to those previews to see if audio data is available. Please note that data may be unavailable for reasons other than diversion, such as internet bandwidth limitations and technical issues.

Prior to August 2016, instead of diverting the entire frequency range of the data, the military diverted only a low frequency band of the data. When this filtering occurred, the file-name was appended with 'LPF' for low-pass filtering. Once filtered, this data remained in separate frequency bands in separate files, so data returned to ONC was labelled 'HPF' for high-pass filtered. See here for more information on the diversion of hydrophone and seismometer data. Data Search and the DataProductDelivery API provide options for filtering out HPF and LPF files, with the "original data" filter in the data diversion mode options (see the Data Product Options section below, one can also select the HPF or LPF files for download). If you happen to download a dataset that has LPF or HPF files and you want to segregate the files, follow these instructions: How to remove LPF and HPF data from an acoustic data set.

For hydrophones located on remote observatories with limited internet bandwidth (Cambridge Bay for instance), live hydrophone audio files may not be available. In this case, any missing audio data is usually stored on site to be retrieved during regular maintenance (yearly or less). Spectrogram data (FFT) files will normally be available near-live as they are much smaller and can be sent over limited connections.

When requested format of audio data already exists in the archive, the search will complete immediately, otherwise some processing will occur to convert the formats. Processing times vary, but in general, expect it would take approximately 15 seconds to process one 5 minute file, so one hour to process a day's worth of data.

Some software applications that employ good error checking may reject some of our wav files for erroneous header information. The cause of this is in the military diversion drivers and is outside of ONC's direct control (they hardcoded these values). Only a select few customized applications will reject the files. Audacity, MATLAB and many others read these files fine. If you get an error like this:
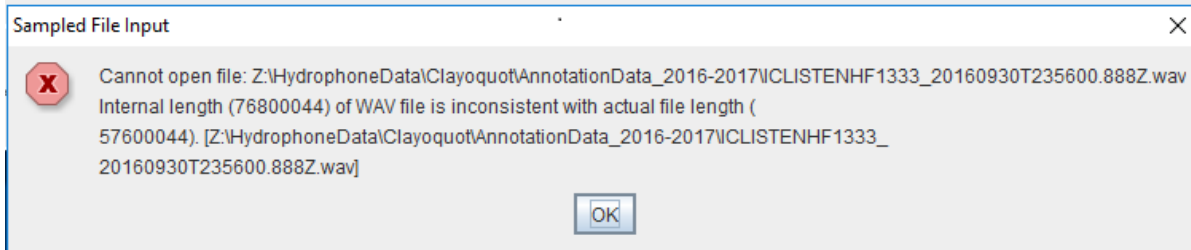
and you are an active MATLAB user, you can fix the wav files with this MATLAB code: oncwavfilefix.m and its dependency: wavchunksizefix.m. If not, please contact us and we'll be happy to help.

See the format sections below the data product options for information on calibration.

Oceans 3.0 API filter: `dataProductCode=AD`

Revision History

1. 20100217: Hydrophone files initially made publicly available
2. 20130123: searches handled by MATLAB search to process data on the fly if needed; all formats made available to all devices

Data Product Options

Hydrophone Channel

For hydrophone data products only (audio and spectrogram data) on the hydrophone array devices only:
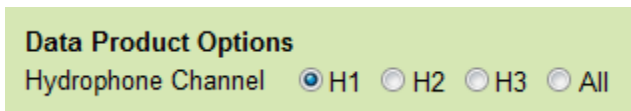
H1

This option will cause the search to return results for hydrophone channel H1 only. The hydrophone arrays consist of multiple hydrophones connected to a single data acquisition computer, which collects the data into single files that have multiple channels (nominally raw hydrophone array files, although other formats can handle multiple channels). Data products may be produced from these files on a per channel basis and returned as specified.

This is the default option.

Oceans 3.0 API filter: `dpo_hydrophoneChannel=H1`

File-name mode field

'H1' is added to the file-name when the hydrophone channel option is set to H1, i.e. IOS3HYDARR02_20111211T152404.000Z-spect-H1.pdf.

H2

This option will cause the search to return results for hydrophone channel H2 only.

Oceans 3.0 API filter: `dpo_hydrophoneChannel=H2`

File-name mode field

'H2' is added to the file-name when the hydrophone channel option is set to H2, i.e. IOS3HYDARR02_20111211T152404.000Z-spect-H2.png.

H3

This option will cause the search to return results for hydrophone channel H3 only.

File-name mode field

'H2' is added to the file-name when the hydrophone channel option is set to H3, i.e. IOS3HYDARR02_20120801T090939.000Z-H3.mp3.

All

This option will cause the search to return results for all available hydrophone channels.

File-name mode field

'H1', 'H2', 'H3', etc are added to the file-name.

Hydrophone Data Diversion Mode

For hydrophone data products only (audio and spectrogram data):

Diversion Mode

Data Product Options

Hydrophone Data Diversion Mode ● Original Data ○ Low Pass Filtered ○ High Pass Filtered ○ All
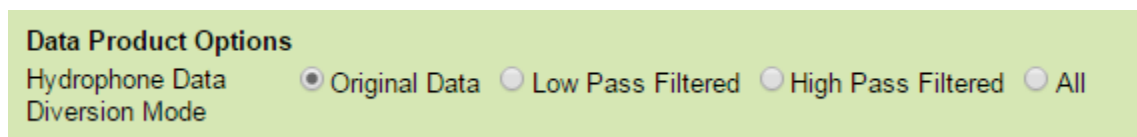
For security reasons, the military occasionally diverts seismic and acoustic data. Over time how this diversion is performed has changed. Currently, when diverted the entire data set is removed. Diverted data is then reviewed by military authorities, if it does not contain sensitive recordings it is returned to the ONC archive.

Standard practice prior to August 2016: instead of diverting the entire data stream, the military diverted only a low frequency band of the data. When this filtering occurred, the remaining data's file-name was appended with 'HPF' for high-pass filtering, while the low-pass data was held for review. Usually that withheld/diverted data was returned, after a delay of 3 days to 2 months; those files are appended with 'LPF' for low-pass filtered. To further confuse matters, sometimes the file-name appending was not complete - half of the data stream was not appended with the LPF or HPF moniker (usually the HPF side), however, our data product software now detects this via time overlaps and handles the other half of the LPF/HPF even if it isn't named so. After 2016, diversions tended to be all or nothing and no low-pass diversion occurred. Recently, the LPF/HPF data splitting has occurred again.

Data diversion is further explained in the data diversion page. Feel free to contact us for support.

Original Data

This option will cause the search to return results for original data only. Files labelled with "-HPF" or "-LPF" are excluded as well as any files that overlap in time with "-HPF" or "-LPF" files. For spectral probability density plots and spectrograms, 'Data Diversion Mode: Original Data' will appear in the plot title.

This is the default option.

Low Pass Filtered

Applies to pre-August 2016 data (with some exceptions). This option will cause the search to return results for diverted data that has been low pass filtered only (only files with "-LPF" in the their file-names). For spectral probability density plots and spectrograms, 'Data Diversion Mode: Low Pass Filtered' will appear in the plot title.

High Pass Filtered

Applies to pre-August 2016 data (with some exceptions). This option will cause the search to return results for diverted data that has been high pass filtered only (only files with "-HPF" in the their file-names). For spectral probability density plots and spectrograms, 'Data Diversion Mode: High Pass Filtered' will appear in the plot title.
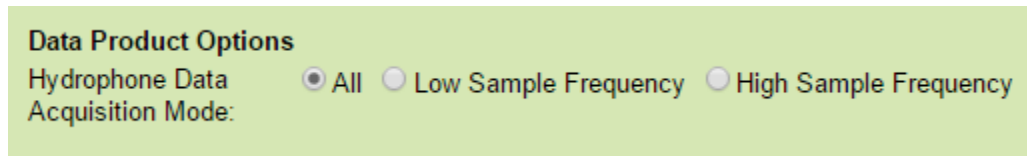
All

This option will cause the search to return results for all data. For spectral probability density plots and spectrograms, 'Data Diversion Mode: High Pass Filtered' will appear in the plot title. This is only way to see data that overlaps in time with files labelled "-LPF" or "-HPF".

File-name mode field

"-LPF" or "-HPF" is added to the file-name when the quality option is set to high or low pass filtered data, i.e. ICLISTENHF1234_20110101T000000Z-HPF. wav. For spectral probability density data products, 'All' may be added to the file-name, as these plots can join LPF, Original and HPF data together into one plot if the spectral frequency bins are the same (data with different frequency content will make addition plots with labels indicating the frequency range). For brevity, 'Original' does not get added to the file-name.

Acquisition Mode

Data Product Options
Hydrophone Data
Acquisition Mode: ● All ○ Low Sample Frequency ○ High Sample Frequency

For hydrophones operating with a duty cycle that includes high and low frequency sample rates (the hydrophones alternate between low and high sample rates periodically, to save battery and memory storage in autonomous deployments). The low sample frequency data will likely have a sample frequency of 16 kHz and the high sample frequency data will likely have a sample frequency greater or equal to then 128 kHz.

Low Sample Frequency

This option will cause the search to return results for the low sample frequency data only (files with "-16KHZ" in their file-names). For spectral probability density plots and spectrograms, "Data Acquisition Mode: Low Frequency" will appear in the plot title.

Oceans 3.0 API filter: `dpo_hydrophoneAcquisitionMode=LF`

High Sample Frequency

This option will cause the search to return results for the high sample frequency data only (files with "-128KHZ" or similar in their file-names). For spectral probability density plots and spectrograms, "Data Acquisition Mode: High Frequency" will appear in the plot title.

Oceans 3.0 API filter: `dpo_hydrophoneAcquisitionMode=HF`

All

This option will cause the search to return results for both the low and high sample frequency data or other mode data. For spectral probability density plots and MAT files, the low and high frequency data will be segregated regardless of option.
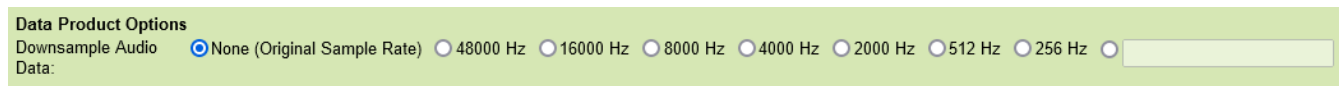
Oceans 3.0 API filter: `dpo_hydrophoneAcquisitionMode=All`

File-name mode field

The sample frequency is added to the file-name for each data acquisition mode option, i.e. ICLISTENHF1234_20110101T000000Z-16KHZ.wav. The Spectrogram_ModeDurationDPO device attribute is populated on devices with a duty cycle, it is used to link the low frequency (LF) and high frequency (HF) acquisition modes with the exact file-name mode modifier string - if this link is not correct, the data acquisition mode option will not properly filter the data products.

Audio Downsampling

For hydrophone audio data products only (MP3, FLAC, WAV formats)

Data Product Options
Downsample Audio
Data: ● None (Original Sample Rate) ○ 48000 Hz ○ 16000 Hz ○ 8000 Hz ○ 4000 Hz ○ 2000 Hz ○ 512 Hz ○ 256 Hz ○ [   ]

Users can specify a target sample frequency for the requested audio data products or the default None option leaves the products unchanged. Users can select from a fixed list of sample frequencies (radio buttons) or specify their sample frequency in the custom input text field, all units are in Hz. When an option other than "None" is selected, the audio data products are downsampled according to standard practice with an anti-aliasing filter (specifically, ffmpeg 's "ar" option with a 32 point FIR Kaiser window having a roll-off of -6 dB at 0.97 of the target sample frequency, additional documentation here and here). If the target sample frequency is the same or greater than the sample frequency in the source file, the user is notified via the search status and no upsampling or resampling takes place. When files are downsampled, it is done on-the-fly. For long search requests, this may take some time and is significantly slower than the "None" option when the requested audio format is archived and available directly. If the requested audio format is not archived and has to be generated from the source format, then downsampling will make the searches return quicker, have smaller file sizes and be quicker to download. See the note at the top of the Audio data product page about what formats are archived. For MP3 formats, downsampling is only applicable at specific rates, which are unknown until an attempt is made to downsample. If downsampling doesn't work, the applicable sample rates will be shown in the search status in the Data Search cart. An example set of MP3 applicable sample rates are:
8000 11025 12000 16000 22050 24000 32000 44100 48000, in Hz.

None (Original Sample Rate)

This option will cause the search to return the data with its original sampling rate.

This is the default option.

Oceans 3.0 API filter: `dpo_audioDownsample=-1`

48000 Hz, 16000 Hz, 8000 Hz, 4000 Hz, 2000 Hz, 512 Hz, 256 Hz

These options will cause the search to return the data at the lowered sampling rate specified. Downsampled files have a modified file name, with the sampling rate added at the end of the file name (ex. '-48000Hz').

Oceans 3.0 API filter: `dpo_audioDownsample=48000`

Oceans 3.0 API filter: `dpo_audioDownsample=16000`

Oceans 3.0 API filter: `dpo_audioDownsample=8000`

Oceans 3.0 API filter: `dpo_audioDownsample=4000`

Oceans 3.0 API filter: `dpo_audioDownsample=2000`

Oceans 3.0 API filter: `dpo_audioDownsample=512`

Oceans 3.0 API filter: `dpo_audioDownsample=256`

Custom

This option will cause the search to return the data at a unique sampling rate specified by the user. Any integer value between 1 and 256000 Hz can be used as input. Including units or any non-numeric characters in this input is not allowed (the red text pops up and the search can't be submitted).

Oceans 3.0 API filter: `dpo_audioDownsample=lowerBnd:1, upperBnd:256000`

File-name mode field

Downsampled files have a modified file name, with the custom sampling rate added at the end of the file name (ex. '-1234Hz').

Audio Format Conversion

For hydrophone audio data products only (MP3, FLAC, WAV formats)

Convert Audio Data From Other Formats: ◯ No (retrieves archived data only - fast - if not downsampling) ● Yes (converts on-the-fly from other formats to fill gaps - slow)

Note: This option is overridden to the "Yes" option if the Audio Downsampling option is any value other than "None".

No

This option will cause the search to return only the data that is in the archive and no on-the-fly format conversion will occur. This saves a tremendous amount of computing time compared to the yes option. The search will normally return in 5-10 seconds, perhaps 30 seconds. The primary storage and source format for hydrophones changed from wav audio files to FLAC audio files on August 10, 2020 with some overlap where both formats were produced and archived, see the note on the audio data product page for more information. Some MP3 files were also archived for 2006 to 2012 (not the source format). The primary benefit of the "no" option is that users can quickly download both source formats (wav and FLAC) with two search requests, one for wav and one for FLAC acquiring all the data quickly without waiting for conversion. Most audio players and software accept both wav and FLAC as input without issue and converting from one format to another has no benefit. The No option will not be in effect if any downsampling is selected, as downsampling is an on-the-fly process, so might as well pull from all sources at that time. The no option is also not in effect for IOS Hydrophone Arrays as the source format for those devices is .hyd raw files which always have to be converted to audio on-the-fly.

This is the default option.

Oceans 3.0 API filter: `dpo_audioFormatConversion=0`

Yes

This option will cause the search to fill in any data files for the requested format that are not already in the archive with files generated on-the-fly, drawing from all source formats (wav, FLAC or hyd (IOS Hydrophone Arrays only)). For example, if a search for wav files is submitted for a year of hydrophone data in 2022 (after the source format was switched to FLAC), all of that data will be converted from the FLAC source files to wav format; this is 105,000 files, which will be about 5 TB on disk and will take about 20 days to process. If you need to convert to a specific format, please limit the search range to one day at a time, as this creates roughly 10 GB of data and takes about 1 hour to process. Making smaller, repeated search requests is best done using the Oceans 3.0 API. It is recommended to test out a small request first before making larger requests.

Oceans 3.0 API filter: `dpo_audioFormatConversion=1`

File-name mode field

No changes.

Format

Hydrophone data is available in **WAV**, **MP3**, and **FLAC** audio files. For normal hydrophones, **WAV** files were the source data format until April 9, 2021, after which FLAC is the source data format. For the IOS hydrophone arrays, HYD files are the source data format. **MP3**s are generated from the source data format (**WAV** or **FLAC**), usually in 128 kilobit quality and were stored in the archive for fast retrieval for most hydrophones until August 10, 2020. After that date, **FLAC** files were generated from the **WAV** source format and archived for fast retrieval. After April 9, 2021, **FLAC** is the source format, stored directly in the archive. All three formats may be generated on-the-fly if they are not pre-generated and archived. On-the-fly generation is relatively slow; users should try to access the archived data formats for best results. Overall, audio / hydrophone data is voluminous and is best handled in small chunks, perhaps via the API: Oceans 3.0 API Home

Some autonomously operated third-party hydrophones only acquired data in **FLAC** format and did not have any formats pre-generated. These files are archived and retrieved without any modification: none of the above data product options apply to **FLAC** files from those hydrophones.

For **WAV, FLAC** and other lossless format data, the files will be accompanied by a calibration file (except for FileDownloadService requests). This file is a text file, comma delimited with one descriptive header row. It is named following the usual standard and ending with '-hydrophoneCalibration.txt'. The dateFrom / dateTo is taken from the calibration date range of the first sensitivity bin attribute. If a search extends over multiple calibrations, multiple files will be produced. The calibrations included are any that overlap with the data time range. A currently applicable or on-going calibration will produce a file named with a dateTo that is midnight tomorrow. Here is an example of the first few lines of a hydrophone calibration text file:

```
#Hydrophone calibration sensitivities. The file contains one header line followed by comma delimited data.
First column is the centre frequency of each frequency bin(Hz). Second column is the sensitivity calibration
for each bin (dB/uPa). Data is from device attributes: http://qaweb2.neptune.uvic.ca/DeviceListing?
DeviceId=1230 . Device attribute HydrophoneSensitivityVectorPart1 last modified: 02-Dec-2015 20:59:52. File
created: 29-Mar-2016 15:36:06.
1, -33.1853
2, -30.6233
3, -29.8957
4, -29.6103
```

Calibrations are stored in the device attributes system, visible in the "additional attributes" tab in device details, for example: https://data.oceannetworks.ca/DeviceListing?DeviceId=23483 Sensitivities are stored in attributes named HydrophoneSensitivityVectorPart# and the frequency bins HydrophoneSensitivityVectorBinsLeadingEdgePart#, broken into parts in order to fit in the system, where # is the part number in order of increasing frequency.

Post-deployment calibrations are also available. The device attribute fields are the same except with a "Post" inserted, e.g. HydrophonePostSensitivityVectorPart1. The format of the calibration text file is the same except that the file-name ends with '-hydrophonePostCalibration.txt' and the header line includes the text 'post-deployment'. Post-deployment calibrations are taken after a device is returned from the field and the frequencies calibrated may only be up to 2000 Hz (lower frequencies tend to show degradation over time, whereas higher frequency responsiveness is more stable). For all calibrations, dateFrom on the device attribute is the time when the calibration was generated. So Data Search requests for post-deployment calibrations include calibrations that occur during the data time range and the first one after the data, if available, so that users have access to the post-deployment calibration.

Oceans 3.0 API filter: `extension={wav,mp3,flac}`

Discussion

To comment on this product, click *Add Comment* below.